Proceedings of Conference on Business Venturing and Entrepreneurship, Spring 2022
April. 16, 2022. pp.123-126

The Impact of Topic Distribution on Review Sentiment: A Comparative Study between South Korea and the U.S.

Mina Cho*
Graduate School of Business, Seoul National University, South Korea, Masters student
Dugmee Hwang**
Graduate School of Business, Seoul National University, South Korea, Masters student
Seongmin Jeon***
College of Business, Gachon University, South Korea, Professor

Abstract

Online reviews offer valuable information to businesses by reflecting consumer experiences about their products and services. Two important aspects of online reviews are first, the topics consumers choose to address and second, the sentiments expressed in their reviews. Building upon previous literature that shows online reviews are context-dependent, we examine the impact of topic distribution on review sentiment in South Korea and the U.S. during pre-and post-pandemic periods. After performing topic modeling on Airbnb app review data, we measure the contribution of each topic on review sentiment using SHAP values. Our results indicate variations in topic distribution trends between 2018 and 2021. Also, the order and magnitude of topics' impact on review sentiment change between pre-and post-pandemic periods for both countries. This study can help businesses to understand how topics and sentiments associated with their products and services changed after pandemic, and also help them identify areas of improvement.

Keywords: Online Review, Text Analysis, Sentiment, Topic Modeling, Pandemic

Ⅰ. Introduction

User-generated online reviews contain a wealth of information, as consumers share their personal experiences of using a product or service. User-generated online reviews are context-dependent. In other words, a change in context or environment is likely to lead to a change in the topic and sentiment contained in the reviews. In the case of the hospitality and lodging industry, COVID-19 has created additional service expectations on top of traditional ones including social distancing, sanitization, and mask-usage(Mehta et al., 2021). Thus, changes in consumers' expectations and evaluations are likely to be reflected in their reviews.

The purpose of this research is to study the effect of topic distribution on the sentiment of online reviews in pre-pandemic(2018-2019) and post-pandemic(2020-2021) contexts. In this study we aim to answer the following research questions: First, do the topic distribution and sentiment of online reviews reflect changes in pre- and post-pandemic user behavior? Second, does the effect of topic distribution on review sentiment differ for pre- and post-pandemic periods?

Ⅱ. Literature Review

2.1. User Generated Content

User generated content(UGC) refers to original and brand-specific content created by customers. UGC includes text, image, video and even voice recordings, and it can be found on social media, website, platforms, etc.

As UGC is a valuable source of information for both

* mina.cho@snu.ac.kr
** dugmee@snu.ac.kr
*** smjeon@gachon.ac.kr

consumers and sellers, previous research have studied how different aspects of online reviews impact consumers and sellers. Consumers use other users' reviews to assist them in making purchase decisions. According to a survey by Podium, almost 93% of consumers agree that they were influenced by online reviews when making purchase decisions(Fullerton, 2017). The survey report also suggests that the influence of online reviews is associated with how much consumers trust the reviews. In the case of sellers, online reviews help sellers to understand what problems their products or services have. Sellers can improve customer satisfaction through correcting such problems. Online reviews help sellers to improve customer relationship by tailoring their products and services to meet consumer needs(Geetha et al., 2017).

Previous research has mainly focused on examining various review characteristics. Liu pointed out that "the literature focuses mainly on review characteristics, such as review length, depth, readability, subjectivity, and sentiments"(Liu et al., 2021). However, relatively little attention has been paid to examining the content of online reviews through text analysis.

2.2. Topic Modeling

Topic modeling is a classification method that is widely used for text analysis. The most popular topic modeling method is Latent Dirichlet allocation(LDA), an unsupervised machine learning technique that identifies latent topic information from large amount of documents(Hong & Davison, 2010).

Past literature has used LDA to extract topics in various contexts. For example, one study combined LDA and hierarchical ward clustering on Airbnb reviews to extract and understand the relationship between topics(Sutherland et al., 2020). LDA has also been used analyze trends during the pandemic periods. Kim et al. used LDA on news articles and tweets to identify topics about the Ebola virus(Kim et al., 2016). In the context of the travel industry, LDA was used to identify consumer needs of the Airline market(Kwon et al., 2021).

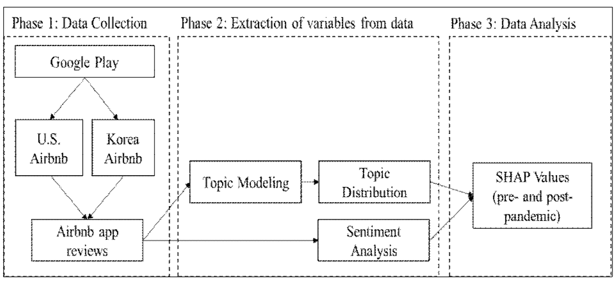
2.3. Sentiment Analysis

Sentiment analysis is used to identify the positive and negative emotions associated with the text. Sentiment is generally categorized into 3 groups: positive, neutral, and negative. Various sentiment analysis tools have been used in previous research. Examples of tools include SentiWordNet(Esuli & Sebastiani, 2006; Ohana & Tierney, 2009), SentiStrength(Guzman et al., 2014; Thelwall, 2017), and Semantic Orientation CALculator(SO-CAL) (Kim et al., 2016; Taboada et al., 2011). Previous literature has used sentiment analysis to examine whether sentiment impacts various characteristics of reviews such as review helpfulness. It was found that negative reviews were deemed less helpful when they expressed intensively negative emotions (M. Lee et al., 2017).

Ⅲ. Research Model

3.1. Hypothesis & Research Framework

This study proposes a framework to evaluate the impact of topic distribution on review sentiment during pre- and post-pandemic periods. The research framework is presented in <Figure 1>. Phase 1 is the data collection stage, where we collected Korean and U.S. Airbnb app reviews by web-scraping from Google Play. In Phase 2, we performed text pre-processing on Korean and English reviews before calculating the optimum number of topics. We then extracted the two main variables used in our study: topic distribution and sentiment score. Finally in Phase 3, we studied the impact of topic distribution on sentiment scores during pre- and post-pandemic periods through SHAP values.



<Figure 1> Proposed Research Framework

Our three hypotheses are as follows:

H1A: The topic distribution of online reviews reflects changes in pre- and post-pandemic user behavior.

H1B: The sentiment of online reviews reflects changes in pre- and post-pandemic user behavior.

H2: The effect of topic distribution on review sentiment is differs for pre- and post-pandemic periods.

3.2. Empirical Context

To answer our research questions, we chose Airbnb as our empirical context. Due to the restrictions on travel and movement caused by COVID-19, hospitality and lodging businesses have been hit hard. As the largest peer-to-peer accommodation sharing platform, Airbnb was no exception. Due to increasing number of cancellations and slowing number of bookings, Airbnb's revenue dropped 72% year-over-year in the second quarter of 2020(Abril, 2020). We chose to analyze Airbnb reviews because they are likely to reflect changes in user behavior between pre-and post-pandemic periods. Due to increased concerns about personal hygiene and cleanliness, topics and sentiments in Airbnb app reviews would reflect distinctive changes in user behavior.

In our work, we perform topic modeling and sentiment analysis on 3485 Korean and 86,909 U.S. Airbnb app reviews collected from Google Play. After text pre-processing Korean and English texts, we extract topic weights and sentiment scores. We then use SHapley additive explanation(SHAP) to measure the impact of topics on sentiments scores.

The outcomes of LDA show that topics have varying distribution trends in pre- and post-pandemic periods. Also, Sentiment analysis shows that sentiment scores for both Korean and English reviews increased between 2018 and 2019 in pre-pandemic periods but decreased with the onset of the pandemic. Finally, our results indicate that the effect of topic distribution on review sentiment differs between pre- and post- pandemic periods. We also compare results for Korean and U.S. Airbnb app reviews and discuss their similarities and differences.

REFERENCE

Abril, D.(2020). *Airbnb's IPO filing reveals huge COVID impact. Fortune*. Retreived from https://fortune.com/2020/11/16/airbnb-ipo-initial-public-offering-coronavirus-impact/.

Esuli, A., & Sebastiani, F.(2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. Proceedings of the 5th International Conference on Language Resources and Evaluation. *LREC 2006*, 417-422.

Fullerton, L.(2017). *Online reviews impact purchasing decisions for over 93% of consumers, report suggests. The Drum*. Retrieved from https://www.thedrum.com/news/2017/03/27/online-reviews-impact-purchasing-decisions-over-93-consumers-report-suggests.

Geetha, M., Singha, P., & Sinha, S.(2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. *Tourism Management*, 61, 43-54.

Guzman, E., Azócar, D., & Li, Y.(2014). Sentiment analysis of commit comments in GitHub: An empirical study. 11th Working Conference on Mining Software Repositories. *MSR 2014-Proceedings*, 352-355.

Hong, L., & Davison, B. D.(2010). Empirical study of topic modeling in Twitter. *SOMA 2010-Proceedings of the 1st Workshop on Social Media Analytics*, 80-88.

Kim, E. H. J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M.(2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, 42(6), 763-781.

Kwon, H. J., Ban, H. J., Jun, J. K., & Kim, H. S.(2021). Topic Modeling and Sentiment Analysis of Online Review for Airlines. *Information*, 12(2), 78.

Lee, M., Jeong, M., & Lee, J.(2017). Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach. *International Journal of Contemporary Hospitality Management*, 29(2), 762-783.

Liu, A. X., Li, Y., & Xu, S. X.(2021). Assessing the unacquainted: Inferred reviewer personality and review helpfulness. *MIS Quarterly: Management Information Systems*, 45(3), 1113-1148.

Mehta, M. P., Kumar, G., & Ramkumar, M.(2021). Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis. *Tourism Recreation Research*. 1-18

Ohana, B., & Tierney, B.(2009). Sentiment classification of reviews using SentiWordNet. 9th. *IT & T Conference.*

Sutherland, I., & Kiatkawsin, K.(2020). Determinants of guest experience in Airbnb: A topic modeling approach using LDA. *Sustainability (Switzerland)*, 12(8), 3402.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M.(2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2), 267-307.

Thelwall, M.(2017). The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength. *In Cyberemotions*, 119-134.